Le lemme de Zorn

Rémi Peyre *

17 avril 2008 †

Résumé

Le but de ce document est de présenter de façon rigoureuse le lemme de Zorn à un public non-spécialiste. La § 1 présente le fameux lemme ; la § 2 le démontre ; la § 3 discute de ses liens avec l'axiome du choix ; et la § 4 en donne quelques applications. Je remercie Jean-Baptiste Guillon dont les discussions sont à l'origine de ce texte.

Table des matières

*L'auteur autorise quiconque à réutiliser le présent document, pourvu que soit respectée la licence *Creative Commons* BY-SA.

†Version retouchée au 2016-10-27.

1 L'objet du délit

1.1 Énoncé du lemme

Avant d'énoncer le lemme de Zorn, nous avons besoin de donner ou rappeler quelques points de vocabulaire :

1.1.1 Définition.

1. Soit X un ensemble. On appelle *ordre strict* sur X une relation binaire $<$ vérifiant les propriétés suivantes :
   — *(Transitivité).* Pour tous $x, y, z \in X$, si $x < y$ et $y < z$, alors $x < z$;
   — *(Antiréflexivité).* Pour tout $x \in X$, $x \not< x$.
2. On appelle *ordre large* — ou plus simplement *ordre* — sur X une relation binaire $\leqslant$ vérifiant les propriétés suivantes :
   — *(Réflexivité).* Pour tout $x \in X$, $x \leqslant x$;
   — *(Transitivité).* Pour tous $x, y, z \in X$, si $x \leqslant y$ et $y \leqslant z$, alors $x \leqslant z$;
   — *(Antisymétrie).* Si $x \neq y$, il est impossible d'avoir simultanément $x \leqslant y$ et $y \leqslant x$.

1.1.2 Remarque. Si $<$ est une relation d'ordre strict, la relation $\leqslant$ définie par $(x \leqslant y) \Leftrightarrow (x < y$ ou $x = y)$ est une relation d'ordre large. Réciproquement, si $\leqslant$ est une relation d'ordre large, la relation $<$ définie par $(x < y) \Leftrightarrow (x \leqslant y$ et $x \neq y)$ est une relation d'ordre strict. En outre, ces deux tranformations sont inverses l'une de l'autre. De ce fait, on associera automatiquement un ordre large à tout ordre strict, et vice-versa.

1.1.3 Remarque. Si E est un ensemble et $X \subset \mathfrak{P}(E)$ un ensemble de parties de E, alors la relation d'inclusion $\subset$ est un ordre large sur X. Réciproquement, il est possible de montrer que tout ensemble ordonné est isomorphe à une structure de ce type.

1.1.4 Définition.

1. Un ordre $\leqslant$ sur X est dit *total* lorsque, pour tous $x \neq y$, on a $x < y$ ou $y < x$.
2. Soit X un ensemble muni d'un ordre $\leqslant$; on dit qu'une partie $C \subset X$ est une *chaine* de X lorsque $\leqslant$ définit un ordre total sur C.

1.1.5 Définition.

1. Soient X un ensemble ordonné, $A \subset X$ une partie de X et $x \in X$ un élément de X. On dit que x est un *majorant* de A lorsque, pour tout $a \in A$, on a $a \leqslant x$.
2. Un ensemble ordonné X est dit *inductif* lorsque chacune de ses chaines admet au moins un majorant.
3. Un élément $m \in X$ est dit *maximal* lorsque, pour tout $x \in X$, $m \not< x$.

1.1.6 Lemme (Zorn)**.** *Tout ensemble inductif admet un élément maximal.*

1.2 Interprétation

Voici une façon d'interpréter intuitivement le lemme de Zorn :

Appelons X l'ensemble incriminé. Une chaine peut être vue comme une progression vers des éléments de plus en plus grands de X. Imaginons donc que nous décidions de poursuivre une telle progression jusqu'à ne plus pouvoir continuer. La chaine ultime que nous obtenons peut à priori se terminer de deux façons : soit il y a en fin de chaine une infinité d'éléments tous plus grands les uns que les autres (comme par exemple à la fin d'un intervalle ouvert), soit il y a un élément de la chaine qui est le dernier (comme par exemple à la fin d'un intervalle fermé). Le premier cas est en fait impossible : en effet s'il se produisait, la chaine ayant un majorant par l'hypothèse, on pourrait ajouter celui-ci à la fin de la chaine, ce qui contredirait l'ultimité de celle-ci. Ainsi la chaine a un plus grand élément m. Mais à nouveau l'ultimité de la chaine interdit qu'aucun $x \in X$ soit strictement supérieur à m, attendu que sinon on pourrait ajouter x en fin de chaine : c'est donc que m est un élément maximal de X.

Dit en une phrase peut-être plus élégante : *« Si au-delà du chemin il y a toujours un lieu, c'est qu'il y a un lieu au bout du chemin »*.

1.2.1 Remarque. Il y a un point délicat dans l'argmentation précédente. Regardons en effet comment nous partons à la recherche de notre élément maximal : on prend un premier élément, puis s'il n'est pas maximal, on prend un second élément strictement plus grand que lui, puis un troisième strictement plus grand que le second, etc. Mais que se passe-t-il si à l'infini on n'a toujours pas trouvé d'élément maximal ? Eh bien... on continue ! La chaine infinie que nous venons former a en effet un majorant qu'on place en bout de chaine, et si celui-ci n'est pas maximal on ajoute encore un élément strictement supérieur, quitte à aller une deuxième fois jusqu'à l'infini, puis une troisième, voire même une infinité de fois... cela n'empêche jamais de continuer ! Comme le chantait Jean-Louis Aubert [1] : *« Au bout, tout au bout, tout au bout, tout au bout du rouleau... Y'a encore du rouleau ! »*. Qu'on puisse aller au-delà de l'infini peut sembler choquant ; pourtant cela ne pose en fait pas de problème

quand le cadre mathématique est bien formalisé. Nous verrons (§ 3) que la subtilité qui tracasse les mathématiciens dans notre argument est de nature tout autre.

1.2.2 Remarque. L'interprétation heuristique que nous avons faite sera exactement le fil directeur de la démonstration du § 2.

1.3 Exemples

1.3.1 Exemple. On considère l'arbre généalogique de l'humanité tel qu'il est au jour d'aujourd'hui, ou plus précisément une vision naïve de cet arbre dans laquelle nous faisons l'hypothèse que l'humanité existe depuis toujours et que tout individu a donc deux parents. Nous supposons également qu'il existe un écart minimal entre l'âge d'un parent et celui de son enfant, mettons de cinq ans (1). Considérons la relation $<$ définie par $x < y \Leftrightarrow$ « x est un ascendant direct de y » ; il est clair $<$ est une relation d'ordre strict. Cet ordre est inductif, car pour toute chaine, autrement dit pour un ensemble d'individus d'une même lignée, le dernier d'entre eux — il y en a nécessairement un — est un descendant direct de tous les autres. D'après le lemme de Zorn, il doit donc y avoir un élément maximal, autrement dit un individu sans descendant : il est facile de se convaincre que c'est effectivement le cas.

1.3.2 Exemple. Nous considérons à nouveau l'arbre généalogique de l'humanité au 17 avril 2008, sauf que la relation $<$ désigne cette fois-ci la *descendance* directe. À nouveau $<$ est une relation d'ordre. Cet ordre a-t-il un élément maximal ? *Non*, car tout individu est un descendant de ses parents au moins ! Si le lemme de Zorn est vrai, cela implique donc que le nouvel ordre n'est *pas* inductif. Et de fait, il ne l'est pas : si nous prenons en effet la chaine de mes ascendants paternels directs (moi, mon père, mon grand-père paternel etc.), on voit que cette chaine remonte jusqu'à la nuit des temps, et par conséquent il est impossible que tous ces individus aient un ancêtre commun, attendu que celui-ci serait nécessairement plus jeune que les plus âgés d'entre eux !

2 Démonstration

Démontrons maintenant le lemme ! À cette fin nous introduisons un peu de vocabulaire personnel :

2.0.1 Définition. Soit X un ensemble ordonné ; nous dirons qu'une chaine $C \subset X$ est *ultime* si elle n'admet aucun majorant *strict* (càd. s'il n'y a aucun $m \in X$ pour lequel $\forall c \in C\ c < x$), et *continuable* dans le cas contraire.

(1). La plus jeune femme ayant jamais donné naissance, Lina Medina, avait 5 ans et 7 mois, cf. [2].

Comme nous l'avions observé au § 1.2, une chaine ultime d'un ensemble inductif a nécessairement un plus grand élément, qui est maximal dans X. Par conséquent, la démonstration du lemme de Zorn se ramène à celle du

2.0.2 Lemme. *Tout ensemble ordonné X admet une chaine ultime.*

Démonstration. Pour toute chaine continuable $C \subset X$, choisissons une fois pour toutes un majorant strict de C, que nous notons $m(C)$. Nous introduisons à nouveau un peu de vocabulaire personnel :

2.0.3 Définition.

1. Pour $C \subset X$ une chaine, on dit qu'un sous-ensemble strict $I \subset C$ en est un *début* si tous les éléments de $C \smallsetminus I$ sont supérieurs à tous les éléments de I. Si en outre $I \neq C$, ce début est dit *strict*.
2. Une chaine $C \subset X$ est dite *bonne* quand, pour tout début strict I de C, $m(I)$ appartient à C et est le plus petit élément de $C \smallsetminus I$.

Nous allons montrer maintenant que si C_1 et C_2 sont deux bonnes chaines de X, alors l'une est un début de l'autre. Considérons les $x \in C_1 \cap C_2$ tels que $\{y \in C_1 \cup C_2\,;\, y < x\} \subset C_1 \cap C_2$; notons C^* l'ensemble qu'ils forment. La définition de C^* assure qu'il s'agit d'un début à la fois des chaines C_1 et C_2 ; par conséquent, si C^* coïncide avec C_1 ou C_2, la chaine en question est un début de l'autre. Il ne reste alors plus qu'à montrer qu'il est impossible que C^* soit distinct à la fois de C_1 et C_2. En effet, dans un tel cas C^* serait un début strict de C_1 et de C_2, et donc, puisque ces chaines sont bonnes, $m(C^*)$ serait dans chacune des C_i, et serait le plus petit élément de chaque $C_i \smallsetminus C^*$. Mais alors, en revenant à la définition de C^*, on pourrait en déduire que $m(C^*) \in C^*$, une contradiction.

Considérons alors l'ensemble $\bar{C} \subset X$ défini comme la réunion de toutes les bonnes chaines de X, i.e. $\bar{C}$ est l'ensemble des points de X qui appartiennent à une bonne chaine au moins. Comme, de deux bonnes chaines, l'une est le début de l'autre, on en déduit que $\bar{C}$ est une chaine [(2)]. Montrons que la chaine $\bar{C}$ est bonne. Soit I un début strict de $\bar{C}$. Soit $j \in \bar{C} \smallsetminus I$; par définition de $\bar{C}$, il existe une bonne chaine de X contenant j : considérons une telle chaine J. On observe alors que $I \subsetneq J$: en effet, pour tout $h \in I$, on sait que h appartient à au moins une bonne chaine H ; et comme H et J sont deux bonnes chaines, l'une est le début de l'autre, et donc dans tous les cas $h \in J$ [(3)]. Puisque $I \subsetneq J$, que I est un début de $\bar{C}$, et que $J \subset \bar{C}$ (par définition de $\bar{C}$), on voit que I est un début strict de J ; et puisque J est une bonne chaine, il s'ensuit que $m(I) \in J \subset \bar{C}$, ce qui prouve bien que la chaine $\bar{C}$ est bonne.

(2). Je détaille l'argument : soient $x, y \in \bar{C}$. x, resp. y, appartient au moins à une bonne chaine C_x, resp. C_y. Mais une de ces deux bonnes chaines est un début de l'autre — disons pour fixer les idées que C_x est une début de C_y —, de sorte que x et y sont deux éléments de la même chaine C_y, et sont donc comparables.

(3). Ici on utilise que $h < j \in J$.

Pour finir, $\bar{C}$ ne peut être continuable, car sinon $\bar{C} \cup \{m(\bar{C})\}$ serait une bonne chaine, et donc, en revenant à la définition de $\bar{C}$, $m(\bar{C})$ appartiendrait à $\bar{C}$, ce qui est absurde. La chaine $\bar{C}$ que nous avons construite est donc ultime. □

3 L'axiome du choix

3.1 Du lien entre lemme de Zorn et axiome du choix

La démonstration ci-dessus vous a-t-elle convaincu de bout en bout ? Tant mieux ! Pourtant, un des outils de la preuve est considéré comme discutable par certains mathématiciens. Cet outil, nous l'avons utilisé au tout début de la preuve du lemme 2.0.2 lorsque nous avons *choisi* simultanément un majorant strict pour *chaque* chaine continuable. Cela n'est justifié que si on s'autorise l'

3.1.1 Axiome (choix)**.** Soit $(X_i)_{i\in I}$ une famille d'ensembles non vides. Alors il existe un *fonction de choix* sur les X_i, c'est-à-dire une application c de I dans la réunion $\bigcup_{i\in I} X_i$ telle que, pour tout $i \in I$, $c(X_i) \in X_i$.

L'axiome du choix (AC) n'est *pas* une conséquence des autres axiomes. Certes, pour tout ensemble $X \neq \varnothing$, les règles du raisonnement nous autorisent à choisir un $x_0 \in X$. L'axiomatique habituelle des ensembles, qu'on appelle ZF, nous permet même de le faire simultanément sur un nombre fini de X_i, comme on le voit par récurrence sur ledit nombre. Mais dès que I est infini, il est impossible de prouver AC à partir de ZF ! Ce fait a été rigoureusement établi par P. Cohen [3] qui a démontré que, si les axiomes de ZF n'engendraient pas de contradiction [(4)], alors aussi bien AC que sa négation pouvaient être ajoutés à ZF sans créer de contradiction logique — on dit que AC est *indécidable* à partir de ZF.

Le lemme de Zorn dépend donc de AC. Mieux, comme nous le démontrerons au § 4.1, il est en fait *équivalent* à AC dans la mesure où on peut démontrer AC à partir de ZF et du lemme de Zorn.

3.2 Pour ou contre l'axiome du choix ?

Je ne sais pas pour vous, mais quand j'ai rencontré l'axiome du choix pour la première fois je me suis demandé pourquoi diable certains le rejetaient. En effet, AC s'accorde me semble-t-il parfaitement à notre vision habituelle des

(4). Comme l'a montré K. Gödel dans son second théorème d'incomplétude [4], même si ZF n'est pas contradictoire, il est impossible de le démontrer à l'aide du seul système ZF — ce résultat restant valable en remplaçant ZF par n'importe quel système axiomatique intéressant.

ensembles ; faire une infinité de choix est peut-être difficile à concevoir, mais dans le monde idéal qu'est celui des mathématiques cela ne choque pas outre mesure — pas plus, en tout cas, que de s'autoriser par exemple à considérer l'ensemble de *toutes* les parties d'un ensemble, comme le fait ZF.

Historiquement, le premier problème soulevé par AC vient de la philosophie générale qui préside à l'établissement de la liste des axiomes de ZF [5, chap. 3]. *Grosso modo*, les axiomes de ZF s'attachent à construire d'une part des ensembles de plus en plus grands à partir d'autres ensembles plus petits, d'autre part des sous-ensembles de ces « gros » ensembles caractérisés par des propriétés du premier ordre. Or AC n'entre pas dans ce paradigme : n'est-il donc pas quelque peu « artificiel » ?

Le second problème — qu'on pourrait dans une certaine mesure connecter au premier — est celui qui préoccupe réellement nombre de mathématiciens. Il réside en ce que AC est utilisé dans toutes sortes de démonstrations *non constructives*, autrement dit des démonstrations grâce auxquelles on prouve l'existence d'un ensemble vérifiant une certaine propriété, *sans* qu'on sache pour autant *caractériser* aucun ensemble vérifiant cette propriété.

3.2.1 Exemple. Soit $\{a, b\}$ un alphabet de deux lettres ; notons X l'ensemble des mots infinis de cet alphabet, par exemple $bbbabbaa\ldots$. On définit sur X une *relation d'équivalence*[(5)] $\sim$, où $x \sim y$ signifie qu'à partir d'une certaine position les mots x et y sont identiques. Par exemple les mots $aabababa\ldots$ (deux a puis une infinité de ba) et $bbbababab\ldots$ (trois b puis une infinité de ab) sont équivalents car ils coïncident à partir de la troisième lettre. Comme $\sim$ est une relation d'équivalence, on peut partitionner X selon ses *classes d'équivalence* : la classe d'équivalence $\bar{x}$ de $x \in X$ est l'ensemble des $y \in X$ équivalents à x, et tous les éléments de $\bar{x}$ ont $\bar{x}$ pour classe d'équivalence, de sorte que tout élément de X appartient à une classe d'équivalence et une seule — la sienne.

D'après AC, il existe donc une fonction de choix sur l'ensemble $X/\!\sim$ des classes d'équivalence de X, ce qui est équivalent à se donner une application $c : X \longrightarrow X$ vérifiant :

- $\forall x \in X \quad c(x) \in \bar{x}$;
- $x \sim y \;\Rightarrow\; c(x) = c(y)$.

Or il est totalement impossible (on peut le démontrer) de *définir* une telle c de manière *explicite*. On voit ainsi le paradoxe : AC nous amène à affirmer l'existence d'objets dont on ne sait donner aucun exemple particulier !

J'aimerais cependant attirer l'attention du lecteur sur un point. L'axiome du choix se distingue des procédures habituelles de choix parce qu'il nous autorise à effectuer simultanément une *infinité* de choix. Néanmoins dans l'exemple que nous venons de voir, on n'a pas l'impression que ce soit de là que vienne le problème, mais plutôt de ce que pour *certaines* classes d'équivalence,

(5). Les propriétés d'une relation d'équivalence sont que $x \sim y$ est équivalent à $y \sim x$, que $x \sim x$ pour tout $x \in X$, et que ($x \sim y$ et $y \sim z$) entraîne $x \sim z$.

on n'a pas de *procédure* canonique permettant d'en choisir un élément [6] ! Or, la simple affirmation « pour toute classe d'équivalence, il existe un élément de cette classe » ne dépend pourtant pas de AC... Et de fait, il s'avère qu'on peut en réalité prouver l'existence d'ensembles impossibles à caractériser dans le cadre du simple système axiomatique ZF (*sans* AC) [6] ! Ainsi, l'argument que AC serait *responsable* des démonstrations d'existence non constructives est quelque peu spécieux : si on souhaite réellement éliminer de telles démonstrations, il faut aller beaucoup plus loin que le simple rejet de AC, en introduisant de nouveaux axiomes en sus... [7]

En conclusion, faut-il accepter l'axiome du choix ? À cette question je réponds à titre personnel que l'axiome du choix est une propriété vraie des « vrais » ensembles — ou, comme dirait Platon [7], des ensembles du monde des idées —, mais que d'autres théories mathématiques, où le concept d'existence aurait une signification différente (par exemple, où il correspondrait à la caractérisabilité), et dont l'étude serait également pertinente, se décrivent par une axiomatique où AC est faux [8]. Par contre, j'ignore s'il est possible de construire une telle théorie en ajoutant certains axiomes à ZF, ou s'il s'agira forcément d'une axiomatique fondamentalement différente.

4 Applications du lemme de Zorn

Nous allons maintenant donner quelques applications du lemme de Zorn. Tous les énoncés dont il sera question ne peuvent être prouvés sans utiliser le lemme de Zorn — ou, ce qui est équivalent, l'axiome du choix. Le lecteur remarquera la puissance du lemme de Zorn qui rend les démonstrations très rapides, ainsi que la grande similitude entre les différents ensembles inductifs introduits.

(6). Comme le disait B. Russell : « Si on a une infinité de chaussettes, on a besoin de l'axiome du choix pour choisir une chaussette de chaque paire ; par contre ce n'est pas le cas pour des chaussures. » (*To choose one sock from each of infinitely many pairs of socks requires the Axiom of Choice, but for shoes the Axiom is not needed*). En effet, un protocole comme « la chaussure gauche » permet de désigner une chaussure particulière de toute paire, alors que, deux chaussettes appariées étant complètement identiques, il est impossible d'en particulariser une *a priori*.

(7). Il serait toutefois abusif de considérer en l'état qu'il n'y a *aucun* lien entre axiome du choix et démonstrations d'existence non constructives, puisque l'argument utilisé dans [6] pour montrer que ZF permet de telles démonstrations utilise précisément le fait que ZF soit *compatible* avec AC.

(8). Une telle axiomatique, l'*intuitionnisme*, a déjà été développée [8]. Mais dans cette théorie, l'existence signifie que l'objet est non seulement caractérisable mais aussi *calculable*, ce qui est un concept beaucoup plus exigeant ! En particulier, l'intuitionnisme récuse le principe du *tiers exclu* qui affirme qu'une proposition est soit vraie, soit fausse, ce qui est tout de même gênant...

4.1 L'axiome du choix

Dans ce paragraphe nous démontrons AC (axiome 3.1.1) à partir du lemme de Zorn.

Démonstration. Soit X un ensemble d'ensembles non vides ; appelons *fonction de choix partielle* sur X la donnée d'un sous-ensemble $Y_c \subset X$ et d'une fonction de choix c sur Y_c. Pour deux fonctions de choix partielles c et c', on dit que c' *prolonge* c si $Y_c \subset Y_{c'}$ et c et c' coïncident sur Y_c.

Notons $c \preccurlyeq c'$ pour « c' prolonge c » ; alors $\preccurlyeq$ est clairement une relation d'ordre. En outre cet ordre est inductif, car si $(c_i)_{i\in I}$ est une chaine sur les fonctions d'ordre partielles, cette chaine a un majorant naturel défini comme sa *réunion*, autrement dit son domaine sera l'ensemble des points de X qui sont dans le domaine d'une c_i au moins, et l'image d'un tel point sera son image par n'importe laquelle des c_i dont il appartient au domaine, laquelle est bien la même quel que soit la c_i choisie grâce à la propriété de chaine.

On applique le lemme de Zorn : il existe alors une fonction de choix partielle maximale $\bar{c}$. Montrons que $\bar{c}$ est une fonction de choix « tout court », autrement dit que $Y_{\bar{c}} = X$: si $X \setminus Y_{\bar{c}} \neq \varnothing$, soient x_1 un de ses élements et ξ_1 un élément de x_1 ; alors la fonction c^* de domaine $Y_{\bar{c}} \cup \{x_1\}$ définie par $c^*(x) = \bar{c}(x)$ pour $x \in Y_{\bar{c}}$ et $c^*(x_1) = \xi_1$ prolonge strictement $\bar{c}$, ce qui contredit la maximalité de $\bar{c}$. C'est donc que $\bar{c}$ est une fonction de choix sur X. □

4.2 Comparaison des cardinaux

Avant d'énoncer le théorème qui fait l'objet de ce paragraphe, nous développons la théorie des cardinaux.

4.2.1 Définition.

1. On dit qu'une application $f : E \longrightarrow F$ est une *injection* quand deux éléments distincts de E ont toujours deux images distinctes par f.
2. On dit qu'une application $f : E \longrightarrow F$ est une *bijection* quand tout élément de F a un unique antécédent dans E par f.

4.2.2 Remarque. Une bijection est donc un cas particulier d'injection ; on peut aussi voir une injection comme une bijection de E dans une partie de F.

4.2.3 Définition. Deux ensembles X et Y sont dits *équipotents*, et on note $|X| = |Y|$, quand il existe une bijection entre X et Y. $|X|$ se lit « *cardinal* de X ». De même, on note $|X| \leqslant |Y|$ pour signifier l'existence d'une injection de X dans Y.

4.2.4 Remarque. Dire que X et Y sont équipotents signifie donc qu'on peut identifier les éléments de X aux éléments de Y *via* une certaine bijection. Il

n'est donc pas surprenant que l'équipotence soit une relation d'équivalence (voir note (5) pour la définition d'une relation d'équivalence), ce qui autorise à parler des cardinaux comme d'entités autonomes — j'entends par là que quand on note $|X| = |Y|$, cela peut *réellement* se voir comme une égalité entre deux objets.

4.2.5 Théorème (Cantor–Bernstein)**.** $\leqslant$ *est une relation d'ordre sur les cardinaux.*

Démonstration. La réflexivité et la transitivité sont claires. La vraie difficulté est de montrer l'antisymétrie, à savoir si X s'injecte dans Y et Y s'injecte dans X, alors il existe une bijection entre X et Y.

Soient donc X et Y deux ensembles, et supposons l'existence de deux injections $f : X \hookrightarrow Y$ et $g : Y \hookrightarrow X$. Pour $x \in X$, resp. $y \in Y$, nous dirons que x, resp. y, est *d'ordre* 0 s'il n'a pas d'antécédent par g, resp. f. Nous dirons qu'il est d'ordre 1 s'il a un antécédent par g, resp. f, et que cet antécédent — unique puisque f et g sont des injections — est d'ordre 0, d'ordre 2 s'il a un antécédent d'ordre 1, etc., et d'ordre infini s'il ne rentre dans aucun des cas d'ordre fini. En d'autres termes, l'ordre de x est le nombre d'antécédents de x qu'on peut remonter par g et f alternativement.

Alors on construit une bijection φ entre X et Y de la façon suivante :
— Si x est d'ordre fini pair, $\varphi(x) = f(x)$;
— Si x est d'ordre fini impair, $\varphi(x)$ est l'unique antécédent de x par g ;
— Si x est d'ordre infini, $\varphi(x) = f(x)$.
Le meilleur moyen de voir que φ est une bijection est de considérer sa *bijection réciproque* $\psi : Y \longrightarrow X$ définie par :
— Si y est d'ordre fini pair, $\psi(y) = g(y)$;
— Si y est d'ordre fini impair, $\psi(y)$ est l'unique antécédent de y par f ;
— Si y est d'ordre infini, $\psi(y)$ est l'unique antécédent de y par f.
On vérifie alors que la composée $\psi \circ \varphi$ (autrement dit l'application consistant à appliquer φ puis ψ) est l'identité de X, et que $\varphi \circ \psi$ est l'identité de Y, ce qui prouve que φ et ψ sont des bijections réciproques l'une de l'autre. □

Quel est le lien entre théorème de Cantor–Bernstein et lemme de Zorn ? Aucun ! Mais le théorème 4.2.5 donne tout son sens au résultat principal de cette partie, à savoir le

4.2.6 Théorème. *L'ordre sur les cardinaux est total, autrement dit, pour X et Y deux ensembles, il y en a au moins un qu'on peut injecter dans l'autre.*

Démonstration. Soient X et Y deux ensembles. Nous appellerons *bijection partielle* entre X et Y la donnée β d'un ensemble de couples $\big((x_i, y_i)\big)_{i\in I}$ où tous les x_i, de même que tous les y_i, sont distincts ; β peut donc également être vue comme une injection d'une partie de X dans Y, ou encore comme une injection d'une partie de Y dans X.

Nous dirons que β' *prolonge* β quand tous les couples de β sont dans β', et nous noterons alors $\beta \preccurlyeq \beta'$. $\preccurlyeq$ est clairement une relation d'ordre, et cet ordre est inductif car une chaine de β est majorée par sa réunion : d'après le lemme de Zorn, il existe donc une bijection partielle maximale $\bar{\beta}$.

Notons $X_{\bar{\beta}}$ l'ensemble des x_i intervenant dans $\bar{\beta}$, resp. $Y_{\bar{\beta}}$ l'ensemble des y_i intervenant dans $\bar{\beta}$. Il est impossible d'avoir simultanément $X \setminus X_{\bar{\beta}} \neq \varnothing$ et $Y \setminus Y_{\bar{\beta}} \neq \varnothing$, car sinon on pourrait ajouter à $\bar{\beta}$ un couple de $(X \setminus X_{\bar{\beta}}) \times (Y \setminus Y_{\bar{\beta}})$, contredisant sa maximalité. C'est donc que, par exemple, $X_{\bar{\beta}} = X$, mais alors β peut être vue comme une injection de X dans Y, ce qui conclut. □

4.3 Le théorème de Hahn–Banach

D'abord, quelques définitions :

4.3.1 Définition. Un *espace vectoriel* est un ensemble non vide d'objets sur lesquels on peut procéder à la multiplication par un nombre réel et à l'addition, et pour lequel les règles de calculs habituelles s'appliquent.

4.3.2 Exemple. L'ensemble des vecteurs du plan [(9)], muni de la multiplication par un scalaire et de la somme vectorielle, est un espace vectoriel.

4.3.3 Définition. Si E est un espace vectoriel, une partie non vide $F \subset E$ stable par multiplication scalaire et par addition est appelée un *sous-espace vectoriel* de E ; F est alors un espace vectoriel pour la restriction des opérations de E.

4.3.4 Exemple. L'ensemble des vecteurs du plan peut être vu comme un sous-espace vectoriel de l'ensemble des vecteurs de l'espace.

4.3.5 Définition. Une *norme* $\|\cdot\|$ sur un espace vectoriel E est une application de E dans $[0, +\infty[$ vérifiant les trois propriétés :

- $\forall x \in E\ \forall \lambda \in \mathbb{R} \quad \|\lambda x\| = |\lambda| \cdot \|x\|$;
- $\forall x, y \in E \quad \|x + y\| \leqslant \|x\| + \|y\|$;
- $x \neq 0 \ \Rightarrow \|x\| > 0$.

Une norme mesure donc en quelque sorte à quel point un élément de E est loin de 0. Un espace vectoriel muni d'une norme est dit *normé*.

4.3.6 Exemple. La norme habituelle sur les vecteurs du plan, i.e. la longueur du déplacement (mesurée, mettons, en mètres) est bien une norme au sens de la définition 4.3.5.

4.3.7 Définition. Pour E un espace vectoriel, on appelle *forme linéaire* sur E une application $f : E \longrightarrow \mathbb{R}$ vérifiant, pour tous $x, y \in E$ et tout $\lambda \in \mathbb{R}$, $f(\lambda x) = \lambda f(x)$ et $f(x + y) = f(x) + f(y)$. Si E est normé, f est en outre dite k-continue, pour $k \geqslant 0$ un réel, si en outre on a, pour tout x, $|f(x)| \leqslant k\|x\|$.

(9). Rappelons qu'un vecteur du plan n'est pas une « flèche » reliant deux points comme on le représente habituellement, mais le *déplacement* correspondant à cette flèche ; par exemple si $ABCD$ est un parallélogramme, $\overrightarrow{AB}$ et $\overrightarrow{DC}$ sont le *même* vecteur.

4.3.8 Exemple. Toujours pour les vecteurs du plan, si $\vec{u}$ est un vecteur donné du plan, le produit scalaire par $\vec{u}$, i.e. l'application $\vec{v} \mapsto \vec{v} \cdot \vec{u}$, est une forme linéaire $\|\vec{u}\|$-continue.

Voici maintenant le théorème central de ce paragraphe :

4.3.9 Théorème (Hahn–Banach)**.** *Soit E un espace vectoriel normé et $F \subset E$ un sous-espace vectoriel de E : F est* ipso facto *normé par la restriction de la norme de E. Soit f une forme vectorielle k-continue sur F, alors il existe une forme linéaire k-continue $\hat{f}$ sur E qui* étend *f, i.e. qui coïncide avec f sur F.*

Démonstration. Appelons *extension partielle* de f toute forme linéaire k-continue g définie sur un sous-espace vectoriel G de E contenant F qui étende f — on notera que f est ainsi une extension partielle d'elle-même. Soit $\preccurlyeq$ la relation « est étendu par » sur les extensions partielles de f ; $\preccurlyeq$ est une relation d'ordre, inductive car toute chaine est majorée par sa réunion (ou par f si la chaine est vide). D'après le lemme de Zorn il y a donc une extension partielle maximale $\bar{g}$. On va montrer que le domaine $\bar{G}$ de $\bar{g}$ est nécessairement E tout entier, ce qui donnera le théorème 4.3.9 en prenant $\hat{f} = \bar{g}$.

Plus précisément, on montre que si $\bar{G} \subsetneq E$, alors on peut étendre strictement $\bar{g}$. Supposons donc $\bar{G} \subsetneq E$. Éliminons tout de suite le cas trivial où $\bar{g} \equiv 0$ sur $\bar{G}$: $\bar{G}$ s'étend alors par la fonction nulle sur E, laquelle est 0-continue. Ce cas éliminé, on considère l'ensemble $K \subset \bar{G}$ défini par $K = \{x \in \bar{G}\,;\, \bar{g}(x) = 1\}$, qui est non vide car, prenant un y vérifiant $\bar{g}(y) \neq 0$, $\frac{1}{\bar{g}(y)} \cdot y \in K$. On pose alors $\Delta = \inf_{x \in K} \|x\|$; comme nous avons fait l'hypothèse que $\bar{g}$ est k-continue, $\Delta \geqslant 1/k$ — k étant nécessairement non nul puisque sinon $g \equiv 0$.

Considérons maintenant un $v \notin \bar{G}$ et étudions la fonction $\rho : \mathbb{R} \longrightarrow [0, +\infty[$ définie par $\rho(\lambda) = \inf_{x \in K} \|v + \lambda x\|$. En appliquant l'inégalité triangulaire dans un triangle dont le premier sommet est 0 et les deux autres sont respectivement dans $v + \lambda \cdot K$ et $v + \mu \cdot K$, on trouve :

$$\forall \mu, \nu \in \mathbb{R} \quad \rho(\mu) + \rho(\nu) \geqslant |\mu - \nu|\Delta. \tag{4.3.1}$$

Posons alors $\lambda_* = \sup_{\lambda \in \mathbb{R}} \{\lambda - \rho(\lambda)/\Delta\}$, resp. $\lambda^* = \inf_{\lambda \in \mathbb{R}} \{\lambda + \rho(\lambda)/\Delta\}$; (4.3.1) nous assure que $\lambda_* \leqslant \lambda^*$. Soit donc $\lambda_0 \in [\lambda_*, \lambda^*]$; pour ce λ_0 on a :

$$\forall \lambda \in \mathbb{R} \quad \rho(\lambda) \geqslant |\lambda - \lambda_0|\Delta \geqslant |\lambda - \lambda_0|/k. \tag{4.3.2}$$

Nous pouvons maintenant définir une extension stricte $\hat{g}$ de $\bar{g}$ dont le domaine sera $\bar{G} \oplus \mathbb{R} \cdot v = \{x \in E\,;\, \exists y \in \bar{G}\ \exists \mu \in \mathbb{R} \quad x = y + \mu \cdot v\}$. Tout $x \in \bar{G} \oplus \mathbb{R}$ s'écrivant de façon unique $\lambda \cdot y + \mu \cdot v$ avec $y \in K$ (exercice !), on peut poser $\hat{g}(x) = |\lambda - \mu\lambda_0|$, pour le λ_0 défini précédemment. La formule (4.3.2) nous assure alors que $\hat{g}$ est bien k-continue, ce qui conclut. □

À titre d'application du théorème de Hahn–Banach, nous allons démontrer la *moyennabilité* de $\mathbb{Z}$:

4.3.10 Définition. On dit qu'une suite de réels $u = (u_n)_{n\in\mathbb{Z}}$ indexée par $\mathbb{Z}$ est *bornée*, et on note $u \in \ell^\infty$, quand les u_n sont tous contenus dans un même intervalle de $\mathbb{R}$. Pour $u \in \ell^\infty$, on note $\|u\|_\infty = \sup_{n\in\mathbb{Z}} |u_n|$; $\|\cdot\|_\infty$ est alors une norme sur l'espace vectoriel ℓ^∞.

Soit τ l'opérateur de *décalage* sur les suites de ℓ^∞, défini par $(\tau u)_n = u_{n-1}$; on dit qu'une application $\mu : \ell^\infty \longrightarrow \mathbb{R}$ est une *moyenne* sur $\mathbb{Z}$ si :

1. μ est une forme linéaire ;
2. L'image par μ d'une suite constante de valeur c est c ;
3. μ est 1-continue ;
4. μ est invariante par τ, i.e. pour toute suite u, $\mu(\tau u) = \mu(u)$.

4.3.11 Remarque. À titre d'exercice, le lecteur pourra démontrer qu'une moyenne sur $\mathbb{Z}$, si elle existe, vérifie nécessairement les propriétés suivantes :

1. Si u est périodique de période L, i.e. $u_{n+L} = u_n$ pour tout n, alors $\mu(u)$ est la moyenne des L termes dont la répétition constitue la suite.
2. Si u_n tend vers une constante c quand $n \longrightarrow \pm\infty$, alors $\mu(u) = c$.
3. Si deux suites u et v vérifient $u_n \leqslant v_n$ pour tout n, alors $\mu(u) \leqslant \mu(v)$.

4.3.12 Théorème. *Il existe une moyenne sur $\mathbb{Z}$.*

Démonstration. Soit Z le sous-espace vectoriel des suites u de ℓ^∞ telles qu'on puisse trouver des suites $v^1, v^2, \ldots v^N \in \ell^\infty$ et des entiers $k_1, k_2, \ldots, k_N$ vérifiant $u = \sum_{i=1}^N \tau^{k_i} v^i$, autrement dit $u_n = \sum_{i=1}^N v^i_{n-k_i}$, qui rendent $\|\sum_{i=1}^N v^i\|_\infty$ arbitrairement petit. Par exemple, la suite alternant des -1 et des 1, ou encore la suite contenant un 1 quelque part et que des 0 ailleurs, sont dans Z.

Nous définissons un espace vectoriel ℓ^∞/Z dont les éléments sont obtenus en *identifiant* les fonctions de ℓ^∞ qui diffèrent d'un élément de Z. Pour une suite $u \in \ell^\infty$, notons $\bar{u}$ l'élément de ℓ^∞/Z correspondant à u ; on pose :

$$\|\bar{u}\|^\tau_\infty = \inf_{v\in Z} \|u + v\|_\infty. \tag{4.3.3}$$

$\|\cdot\|^\tau_\infty$ est alors une norme sur l'espace vectoriel ℓ^∞/Z.

Notant $\mathbf{1}$ la suite constante de valeur 1, nous voulons maintenant montrer que $\|\bar{\mathbf{1}}\|^\tau_\infty = 1$. Pour cela, on se donne des suites $v^i \in \ell^\infty$, $i = 1, \ldots, N$ et des $k_i \in \mathbb{Z}$ correspondants, avec $\sum_{i=1}^N \tau^{k_i} v^i = \mathbf{1}$, et on va montrer que $\|\sum_{i=1}^N v^i\|_\infty \geqslant 1$. Notons $M = \max_{1\leqslant i\leqslant N} \|v^i\|_\infty$ et $K = \max_{1\leqslant i\leqslant N} |k_i|$. Soit $L \in \mathbb{N}$, moralement très grand. $v^i_{-L-k_i} + v^i_{-L+1-k_i} + \ldots + v^i_{L-k_i}$ différant de $v^i_{-L} + v^i_{-L+1} + \ldots + v^i_L$ par au plus $2K$ termes majorés chacun par M, ces sommes diffèrent au plus de $2KM$. En sommant sur i, on trouve alors que :

$$\Big|(2L+1) - \sum_{n=-L}^{L} \big(\sum v^i\big)_n\Big| \leqslant 2KMN, \tag{4.3.4}$$

en particulier $\sum_{n=-L}^{L}(\sum v^i)_n \geqslant 2L+1-2KMN$, et par conséquent un des termes de $\sum v^i$ est au moins égal à $1-2KMN/(2L+1)$. En faisant tendre L vers l'infini, on obtient que $\sup_{n\in\mathbb{Z}} \sum v_n^i \geqslant 1$, d'où $\|\sum v^i\|_\infty \geqslant 1$. Ainsi la norme de $\bar{1}$ est au moins égale à 1, donc c'est 1.

La suite est facile : soit $m : \mathbb{R}\cdot\bar{1} \longrightarrow \mathbb{R}$ la forme linéaire qui, à la classe dans ℓ^∞/Z d'une suite constante, associe la valeur prise par cette suite. D'après ce que nous venons de voir, m est bien définie et elle est 1-continue sur $\mathbb{R}\cdot\bar{1}$; le théorème 4.3.9 nous permet donc de l'étendre en une forme linéaire $\hat{m}$ 1-continue sur ℓ^∞/Z, et alors $u \longrightarrow \hat{m}(\bar{u})$ est une moyenne sur $\mathbb{Z}$. □

4.4 Le théorème de Zermelo

Enfin, il ne serait pas honnête de terminer ce cours sans parler du théorème de Zermelo. Cet énoncé, équivalent à l'axiome du choix et donc au lemme de Zorn, tient une place centrale en théorie des ensembles ; s'il est peut-être un peu plus délicat à appréhender que le lemme 1.1.6, cela tient surtout de ce qu'il repose sur le concept de *bon ordre*, lequel ne fait pas partie de l'outillage mathématique de licence. Nous allons donc dans ce paragraphe définir les bons ordres, expliquer ce qui fait leur importance, et démontrer le théorème de Zermelo. En lisant ce qui suit, vous constaterez que le concept de bon ordre était en fait déjà caché dans notre démonstration du lemme de Zorn (§ 2).

4.4.1 Définition. Soit $\leqslant$ un ordre sur un ensemble X. On dit que $\leqslant$ est *bon* si toute partie non vide de $A \subset X$ a un *minimum* pour $\leqslant$, c.-à-d. s'il existe $a_0 \in A$ vérfiant $a_0 \leqslant a$ pour tout $a \in A$ [(10)].

4.4.2 Remarque. Un bon ordre est forcément total, comme on le voit en prenant pour A une paire $\{a, b\}$.

Dès qu'un ensemble est bien ordonné, on dispose d'un critère pour choisir un élément d'une de ses sous-parties non vides — attendu qu'un minimum est nécessairement unique —, ce qui montre l'existence d'un lien entre bons ordres et axiome du choix. Les bons ordres ont une propriété fascinante, que nous donnons ici sans démonstration [(11)], et qui explique pourquoi c'est à partir d'eux qu'on a choisi de construire toute la théorie des ensembles [5] :

4.4.3 Proposition. *Soient X et Y deux ensembles bien ordonnés. Alors un des deux ensembles est isomorphe à un début [(12)] de l'autre, càd. qu'il existe une bijection entre le premier ensemble et le début du second ensemble qui préserve l'ordre.*

(10). Attention à ne pas confondre le concept de maximum (ou de minimum) avec celui d'élément maximal (ou minimal), cf. déf. 1.1.5-3.
(11). La démonstration est en fait essentiellement la même que lorsque nous avons expliqué, p. 5 du présent document, que, étant données deux bonnes chaines, il y en a une qui est le début de l'autre.
(12). Cf. définition 2.0.3-1.

La proposition 4.4.3, en clair, signifie qu'il existe une structure de bon ordre *universelle*, dont les débuts correspondent exactement aux structures particulières de bon ordre [13].

4.4.4 Théorème (Zermelo)**.** *Tout ensemble peut être doté d'un bon ordre.*

Démonstration. Soit X un ensemble ; appelons *bon ordre partiel* sur X la donnée d'un $Y \subset X$ et d'un bon ordre sur Y. Soit $\preccurlyeq$ la relation « être prolongé par » sur les bons ordres partiels de X ; c'est un ordre, inductif car toute chaine est majorée par sa réunion. Considérons alors, *via* le lemme de Zorn, un élément maximal $(\bar{Y}, \leqslant)$ pour $\preccurlyeq$. Si $\bar{Y} \subsetneq X$, prenant un $a \in X \smallsetminus \bar{Y}$, on peut alors définir un bon ordre sur $\bar{Y} \cup \{a\}$ en conservant le même ordre pour les éléments de $\bar{Y}$ et en posant $y \leqslant a$ pour tout $y \in \bar{Y}$. Mais alors ce bon ordre prolonge celui sur $\bar{Y}$, ce qui est exclu par maximalité ; c'est donc que $\bar{Y} = X$ et ainsi le bon ordre sur $\bar{Y}$ est un bon ordre sur X. □

4.4.5 Remarque. Le théorème de Zermelo, comme nous l'avons dit ci-dessus, est en fait équivalent à l'axiome du choix. Et de fait, il est impossible par exemple de construire explicitement un bon ordre sur un ensemble aussi simple que $\mathbb{R}$! Ce qui inspira sans doute ce trait d'esprit au mathématicien Jerry Bona : *« L'axiome du choix est trivialement vrai ; le théorème de Zermelo est trivialement faux ; et que pouvons-nous dire du lemme de Zorn ? [14] »*. Ce qui sera notre conclusion.

Références

[1] J.-L. AUBERT : Le Bout du Rouleau, 1989.

[2] La plus jeune mère du monde. *La Presse médicale*, n° 47, 13 mai 1939.

[3] P. COHEN : The Independence of the Continuum Hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 50(6): 1143–1148, 1963.

[4] K. GÖDEL : Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931.

[5] P. DEHORNOY : Logique et théorie des ensembles. Notes de cours ENS/ FIMFA, disponibles à l'adresse `http://www.math.unicaen.fr/~dehornoy/surveys.html`, 2007.

[6] `http://mathoverflow.net/questions/123608`.

[7] PLATON : *La République*. LGF, 1995.

(13). La structure en elle-même n'est pas un bon ordre, car elle contient « trop » d'éléments pour qu'on puisse y attacher un ensemble, ce qui évite de tomber dans des paradoxes.
(14). En anglais : *The axiom of choice is obviously true, the well-ordering principle obviously false, and who can tell about Zorn's lemma?*

[8] D. van DALEN : Intuitionistic Logic. *In The Blackwell Guide to Philosophical Logic*. Blackwell, 2001. sous la direction de L. Goble.